

## High-Performance Computing at Los Alamos announces milestone for key/value middleware

May 26, 2014



### Billion inserts-per-second data milestone reached for supercomputing tool

LOS ALAMOS, N.M., May 29, 2014—At Los Alamos, a supercomputer epicenter where “big data set” really means something, a data middleware project has achieved a milestone for specialized information organization and storage. The Multi-dimensional Hashed Indexed Middleware (MDHIM) project at Los Alamos National Laboratory recently achieved 1,782,105,749 key/value inserts per second into a globally-ordered key space on Los Alamos National Laboratory’s Moonlight supercomputer.

“In the current highly parallel computing world, the need for scalability has forced the world away from fully transactional databases and back to the loosened semantics of key value stores,” says Gary Grider, High Performance Computing division leader at Los Alamos.

Computer simulations overall are scaling to higher parallel-processor counts, simulating finer physical scales or more complex physical interactions. As they do so, the simulations produce ever-larger data sets that must be analyzed to yield the insights scientists need.

“This milestone was achieved by a combination of good software design and refined algorithms. Our code is available on Github and we encourage others to build upon it,” said Hugh Greenberg, project leader and lead developer of the MDHIM project.

Traditionally, much data analysis has been visual; data are turned into images or movies. Statistical analysis generally occurs over the entire data set. But more detailed analysis on entire data sets is becoming untenable due to the resources required to move/search/analyze all the data at once. The ability to identify, retrieve, and analyze smaller subsets of data within the multidimensional whole would make detailed analysis much more practical. In order to do achieve this, it becomes essential to find strategies for managing these multiple dimensions of simulation data.

The MDHIM project aims to create a middle-ground framework between fully relational databases and distributed but completely local constructs like “map/reduce.” MDHIM allows applications to take advantage of the mechanisms provided by a parallel key-value store: storing data in global multi-dimensional order and sub-setting of massive data in multiple dimensions as well as the functions of a distributed hash table with simple but massively parallel lookups.

Records are sorted globally in whichever number of ways an application chooses. Applications can choose to implement, via the MDHIM library, anything from a shared-nothing map/reduce-style functionality to deeply indexed data with rich information about statistical distributions within all keys. This allows global statistical analysis and retrieval of relevant data subsets for further analysis.

MDHIM is designed to represent petabytes of data with mega- to gigabytes of representation data, utilizing the natural advantages of HPC interconnects—low latency, high bandwidth, and collective-friendliness—to scale key/value service to millions of cores, implying a need for billions of inserts per second.

In this sample scaling run, MDHIM ran as an MPI library on 3360 processors within 280 nodes of the 308-node Moonlight system in demonstrating nearly two billion inserts per second.

MDHIM is a framework on which an application can run thousands of copies of existing key value stores, in multiple programming environments, exploiting the capabilities of an extreme scale computing system. MDHIM, which is sponsored by the U.S. Department of Defense, is being used extensively within the Storage and I/O portion of the DOE FastForward project, which has as its objective “to initiate partnerships with multiple companies to accelerate the R&D of critical component technologies needed for extreme-scale computing.”